



eCOMMONS

Loyola University Chicago  
Loyola eCommons

---

Master's Theses

Theses and Dissertations

---

1995

## Do the Effects of Child Psychotherapy Vary Over Time?

James Kelly Cotten

Follow this and additional works at: [https://ecommons.luc.edu/luc\\_theses](https://ecommons.luc.edu/luc_theses)

---

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-sa/3.0/).  
Copyright © 1995 James Kelly Cotten

LOYOLA UNIVERSITY OF CHICAGO

DO THE EFFECTS OF CHILD PSYCHOTHERAPY VARY OVER TIME?

A THESIS SUBMITTED TO  
THE FACULTY OF THE GRADUATE SCHOOL  
IN CANDIDACY FOR THE DEGREE OF  
MASTER OF ARTS

DEPARTMENT OF PSYCHOLOGY

BY  
JAMES KELLY COTTEN

CHICAGO, ILLINOIS

MAY, 1995

Copyright by J. Kelly Cotten, 1995

All rights reserved.

## ACKNOWLEDGEMENTS

I would like to thank my colleagues Anne Wells and Kate McGlinchey, as their assistance and support on this project was crucial to its completion. In addition, I wish to thank the members of my thesis committee, Dr. Joseph Durlak and Dr. Fred Bryant, for their guidance and support of this project. Finally, I am most grateful to my family, whose ongoing love and support enabled my pursuit of this research.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	iii
LIST OF TABLES . . . . .	v
Chapter	
I. INTRODUCTION AND REVIEW OF RELATED LITERATURE . .	1
II. THESIS OVERVIEW AND HYPOTHESES . . . . .	12
III. METHOD . . . . .	24
IV. EXPERIMENT RESULTS . . . . .	36
V. EXPERIMENT DISCUSSION . . . . .	54
Appendix	
A. JOURNALS SEARCHED FOR ORIGINAL RESEARCH REPORTS .	62
B. CODING FORM . . . . .	63
C. EFFECT SIZE CALCULATION PROCEDURES . . . . .	69
D. REFERENCES OF THE STUDY SAMPLE . . . . .	71
REFERENCES . . . . .	84
VITA . . . . .	90

## LIST OF TABLES

Table	Page
1. Selected Characteristics of Studies . . . . .	37
2. Homogeneity Test of Moderator Variables . . . . .	45
3. Weighted Regression Predicting FU ES's . . . . .	48

## CHAPTER I

### INTRODUCTION AND REVIEW OF RELATED LITERATURE

In psychotherapy research, the need for assessing both the initial effects of a treatment (post-treatment effects) and the sustainability of these effects after treatment termination (follow-up effects) has long been recognized (see Raimy, 1952). As pointed out by several psychotherapy researchers (e.g. Nicholson & Berman, 1983; Achenbach, 1978), posttreatment (PT) effects may not reflect or predict the long-term effects of a treatment. Thus, PT analyses fail to address four important questions: (a) which initial treatment effects deteriorate after the termination of treatment? (b) which treatment effects emerge or intensify with the passage of time? (c) which treatment effects show no variation over time? and (d) do follow-up (FU) effects vary as a function of client and/or treatment variables, such as type of problem or type of treatment?

#### Empirical Status of FU Effects

As the number of treatment outcome studies has increased, so have efforts to integrate the results of these studies (e.g., Wright, Moelis, & Poolack, 1976; Smith, Glass, & Miller, 1980; Nicholson & Berman, 1983). The most thorough synthesis of FU (FU) effects to date is Nicholson and Berman's (1983) meta-analysis of multiple treatments for

adults with neurotic disorders.

### FU effects of psychotherapy with adults

Nicholson and Berman (1983) identified 67 studies involving comparisons of either individual or marital therapy for neurotic disorders in adults. These studies were analyzed in three separate ways to address the following questions: (a) does a patient's status at PT accurately predict his or her status at FU? (b) do differences between treatment groups at PT accurately reflect differences at FU? and (c) does the status of a treatment group remain stable from PT to FU?

In addressing the first question, Nicholson and Berman (1983) analyzed studies that either reported a correlation between PT and FU effects for each individual participant in the study, or provided the information necessary to derive this statistic. Of the 67 total published reports cited however, only five provided this information. Thus, Nicholson and Berman's (1983) first analysis was limited to the 68 correlations reported by these five studies. A single correlation was derived for each of the studies by combining the correlations of each individual participant, and then these scores were combined to produce a single mean correlation across the five studies. The results indicate that a patient's standing at PT is a strong predictor of his or her status at FU (weighted  $r(68) = .66$ ,  $p < .001$ ). However, no information was provided regarding the range and mean of



FU interval lengths. As a result, the extent to which the PT status of a patient predicts that patient's status at any particular later point in time is not clear. Furthermore, as noted by the researchers, the small number of studies contributing to this outcome renders it equivocal.

In their second analysis, Nicholson and Berman (1983) compared the relative standing of treatment groups at PT with their relative standing at FU to assess their similarity. Each group comparison ( $N = 78$ ) contributed one PT and one FU effect size (treatment group mean minus control group mean divided by control group's standard deviation). The mean number of months from PT to FU assessment was 8.6, and FU intervals ranged from 1 to 114 months.

The mean effect size at FU differed slightly, though not significantly, from that at PT (.55 versus .68, respectively), and both differed reliably from zero. Subsequent analyses suggested that the slight discrepancies between PT and FU assessments were due to comparisons in which attrition from PT to FU was high, participants at FU were not representative of participants at PT, or treatment groups received inequivalent amounts of additional therapy during the FU interval.

In order to reduce the effects of these nuisance variables, comparisons ( $N = 31$ ) were eliminated if the overall attrition rate was greater than 30%, if the difference in attrition between groups was equal to or

greater than 20%, if participants at FU differed significantly on "some characteristic" (not specified by Nicholson and Berman, 1983) from those at PT, or if the "difference between groups in additional therapy was 20% or more or was reported to be significant" (p. 266). All subsequent analyses were limited to the remaining comparisons ( $N = 47$ ). Unfortunately, all of these procedures were not described clearly enough to replicate, and it is possible that they may have introduced a sampling bias in subsequent analysis. Furthermore, Nicholson and Berman (1983) did not report how this procedure affected the mean and range of FU intervals of the remaining comparisons. Nonetheless, when comparisons meeting these criteria were eliminated, the effect sizes at PT and FU were found to be identical (.70 versus .70 respectively).

Nicholson and Berman (1983) then compared the relative PT and FU effect sizes for specific types of therapy and specific diagnoses or problems. No significant differences in effect sizes were found among any treatment types between PT and FU. Significant differences did emerge as a function of problems, however. Effect sizes for phobias (PT  $ES = .89$ ; FU  $ES = .74$ ) and social problems (PT  $ES = .81$ ; FU  $ES = .96$ ) change significantly from PT to FU. No significant differences across other problems were found.

To determine if differences in the length of the FU interval contributed to these findings, a correlation was

calculated between the length of FU in months and the difference between PT and FU effect sizes. No significant relationship was found in this analysis. The possibility that a nonlinear relationship might exist between these variables was assessed using regression analysis, and again no significant results occurred. These results suggest that the effects realized by treatment groups at both PT and FU are similar across distinct treatments and diagnoses with two exceptions: phobias and social problems. Furthermore, effects do not vary as a function of the length of the FU interval.

While the second part of the Nicholson and Berman (1983) meta-analysis assessed differences that occurred between treatment and control groups during the FU interval, the third phase assessed whether differences occurred within treatment groups. That is, effect sizes were computed from the difference between the means of the treatment group at PT and FU. This analysis revealed no significant change during the FU interval across either different treatments or problems. An analysis of change across different outcome measures likewise found no reliable differences. Again, neither a correlational nor a regression analysis suggested any significant linear or nonlinear relationship between effect size and length of the FU interval. These findings suggest that effects realized by a treatment group at PT are durable over time across different treatments, diagnoses,

and outcome measures.

Although the Nicholson and Berman (1983) meta-analysis provides encouraging evidence for the sustainability of gains achieved through different forms of psychotherapy in adult populations with neurotic disorders, the results are not unequivocal. The statistical power of this meta-analysis to detect significant variance among effect sizes is decreased by two particular characteristics: (a) the equal weighting of effect sizes derived from studies with different sample sizes, and (b) the failure to investigate the influence of several variables which differ across studies. Furthermore, the exclusion of unpublished comparisons in the meta-analysis limits the representativeness of the findings. While this is not an exhaustive list of weaknesses inherent to this meta-analysis, these problems do represent its greatest limitations, and warrant further explanation.

In failing to weight effect sizes according to the sample size from which they were derived, a study with a small sample size, say 10, counts equally with a study with a large sample size, say 1000. As a small sample study contains more error variance than a large sample study, failure to weight the studies accordingly introduces a large and unnecessary degree of error variance to the overall effect size estimates. Thus, effect sizes should be weighted relative to the samples from which they are derived

in order to produce a more reliable population estimate (Hedges & Olkin 1985; Bryant, 1986). Without such weighting procedures, systematic variation among effect sizes may be masked by the influence of less reliable estimates. Unfortunately, Nicholson and Berman (1983) employed no such weighting procedures. Thus, it is reasonable to question whether differences across treatments and problem types might have been found in Nicholson and Berman (1983) had such weighting procedures been used. In addition to these weighting procedures, a more thorough analysis of the influence of the various treatment, therapist, patient, and methodological characteristics represented in the sample of studies may have turned up a systematic variation in effect sizes that is not apparent in the current presentation of results. Although several important factors were considered, such as treatment type, diagnosis, outcome measure, and length of FU interval, the examination of other variables, which will be discussed in the next section, could have increased the ability of the meta-analysis to detect sources of systematic variation among effect size estimates (Hedges & Olkin, 1985; Bryant, 1986). Again, this concern suggests that the results of the Nicholson and Berman (1983) meta-analysis may have failed to detect real differences among FU effects.

Finally, the external validity of Nicholson and Berman (1983) is threatened by its inclusion of only published

studies. The exclusion of unpublished studies has been found to artificially inflate effect sizes (Smith, 1980). Thus, it is questionable whether the effect size estimates found by Nicholson and Berman (1983) are generalizable to all similar treatment situations.

Despite the limitations of the Nicholson and Berman (1983) meta-analysis, the results are encouraging. While there may be individual cases to the contrary, this meta-analysis supports the notion that practitioners, theoreticians, and policymakers can generally expect that changes occurring in those adults who are treated psychotherapeutically for neurotic disorders will not immediately diminish following treatment termination. Yet it should be underscored that the findings of the Nicholson and Berman (1983) meta-analysis are limited to adult populations.

#### FU effects of psychotherapy with children

As pointed out recently by Kazdin (1993) and Durlak, Fuhrman, and Lampman (1991), currently there is insufficient evidence to draw any conclusions about the FU effects of psychotherapy with child populations. Although individual studies of the FU effects of child psychotherapy are substantial in number, these studies have often produced inconsistent findings. For example, in FU studies of the treatment of social relation problems in children, Gottman (1977) reported that therapeutic effects dissipated over time, Jakibchuck and Smeriglio (1976) reported that the

effects increase, and Weinrott, Corson, and Wilchesky (1979) reported that the effects remain stable after treatment termination.

Such inconsistencies in FU studies of child treatment are typified by the review of Wright et al., (1976), who qualitatively evaluated a small sample of six studies that offered comparisons between PT and FU assessments of children receiving individual psychotherapy. They found little agreement among the studies regarding FU effects. For example, three of the studies (Lehrman, Sirluck, Black, & Glick, 1949; Seeman, Barry, & Ellinwood, 1964; Heinicke, 1969) found that the benefits of the treatments increased during the FU interval, while the other three found either a decrease in benefits (Love, Kaswan, Bugental, 1972) or no difference (Dorfman, 1958; Miller, Barrett, Hampe, & Noble, 1972). Differences at FU were hypothesized to be due to differences in the number of therapy sessions, as the three studies involving 30 or more sessions showed an increment at FU while those with fewer demonstrated either no change or a decrement. However, in the absence of both a larger sample of studies and a statistically-driven synthesis, any conclusion based on these observations is premature. Currently, there are over 100 evaluations of the FU effects of child psychotherapy present in the literature, and there is no reason to believe that the sample utilized in the Wright et al. (1976), review is representative of this

extensive data base. Furthermore, the qualitative nature of the Wright et al. (1976), review does not have the power of more current, statistically-driven methods of research synthesis to either yield reliable estimates of treatment effects, or identify independent variables that might influence FU effects (Hedges & Olkin, 1985).

Later reviews of child psychotherapy research have employed the more rigorous approach of meta-analysis, yet have focused primarily on PT effects (Casey & Berman, 1985; Weisz, Weiss, Alicke, & Klotz, 1987; Durlak, Lampman, & Wells, 1990). Of these meta-analyses, only Weisz et al. (1987) reported on FU effects. Among studies including both PT and FU evaluations, Weisz et al. (1987) reported identical means for PT and FU effects (0.93). Unfortunately, the number of studies included in this analysis was not provided. Furthermore, the standard deviation was not reported, and as previously mentioned, a single pooled effect size estimate cannot be trusted unless one is willing to assume that all studies included in the analysis are drawn from the same population. In the Weisz et al. (1987) meta-analysis, a great deal of heterogeneity exists among characteristics of the sample of studies. For example, 21 different treatments and eight different target problems are represented within the sample (Weisz et al., 1987). Thus, the single effect size averaged for all studies reported may be an inappropriate measure, as it fails to account for the



variance among estimates that one would expect to find in such a diverse sample.

## CHAPTER II

### THESIS OVERVIEW AND HYPOTHESES

As noted most recently by Kazdin (1993), assessing the long-term impact of child psychotherapy is exceedingly important in light of the suffering endured by children and their families due to childhood emotional and behavioral problems. Many difficulties experienced by children, such as poor social relations, depression, and conduct disorders, tend to persist into adulthood, accentuating the need for effective childhood interventions (Robins & Rutter, 1990; Weiss & Hechtman, 1986). It is critical to provide caregivers with the information they need to select treatments which will affect lasting improvement in the individuals whom they treat.

The goal of the present investigation is to evaluate through meta-analysis the four previously mentioned issues regarding the FU effects of child psychotherapy: (a) which treatment effects deteriorate after treatment termination? (b) which treatment effects emerge or increase over time? (c) which treatments have effects which remain stable with the passage of time? and (d) what factors moderate the long-term effects of treatment? The scope of problems and treatments included in this assessment will be wide with

the following exceptions: the meta-analysis will not include studies of children with problems of drug use (including smoking), or studies of peer counseling or family therapy.

In prelude to addressing the above issues, a discussion of possible moderating factors and the meta-analytic technique will aid the reader's understanding of subsequent analyses.

### Moderating Variables

Child treatment efficacy studies are rarely identical across subject, treatment, therapist, and study characteristics (e.g., problem severity, treatment type, therapist experience, subject drop-out). Therefore, the FU effects of child psychotherapy may vary as a function of each of these factors. While the following list of possible moderating variables is not exhaustive, it represents the features most often implicated and reported by researchers in our base of studies. Each of these factors will be assessed in this review.

### Type of Problem

In a meta-analysis of treatment efficacy with children, Casey and Berman (1985) found that most reported problems were treated with comparable effectiveness with one exception: the treatment of social relation difficulties was not met with equal success. In a similar review, Weisz et al. (1987) found no reliable difference in PT effects across

problem types. However, all emotional and behavioral problems within the scope of this review may not be equally amenable to lasting change, and this possibility will be explored.

#### Severity of problem

Although problem severity is directly related to problem type in most cases, this is not always so. It is possible for children to suffer differing degrees of impairment due to the same problem. Several researchers have pointed to problem severity as a potential source of variance in treatment outcomes (Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz, Weiss, & Donenberg, 1992). Although this factor has not been addressed independently of problem type in previous reviews (e.g. Casey & Berman, 1985; Weisz et al., 1987), it will be in the present analysis.

#### Age of treated children

Assessing the influence of age on FU effects is actually an attempt to get at the impact of developmental stages. The call for information regarding the moderating influence of developmental stage on treatment effects has been echoed in recent reviews (Kazdin, Bass, Ayers, & Rodgers, 1990; Kazdin, 1993). The importance of assessing the influence of stage of development has been demonstrated by Durlak, Fuhrman, and Lampman (1991), who reported that children's cognitive developmental stage was the primary moderating factor in the outcome of cognitive/behavior

therapy. Unfortunately, few studies report the developmental stage of their subjects, so age must be used as an approximation. Weisz et al. (1987) found that when studies were divided into the broad categories of children (ages 4-12) and adolescents (ages 13-18), treatment effects for the former group were significantly larger than for the latter. However, it is not clear if a distinction of this nature exists in FU effects.

#### Gender of treated children

Previous research suggests that male children benefit to a lesser degree from therapy than do females (Casey & Berman, 1985; Weisz et al., 1987). Thus, it is reasonable to question whether the benefits realized by males are as sustainable as those in females.

#### Ethnicity of treated children

Currently, there is a deficiency in the integration of information regarding treatment effects on children from minority populations. Ethnicity of subjects has been largely ignored in previous meta-analyses of treatment outcome studies with children (i.e. Casey & Berman, 1985; Weisz et al., 1987). As a result, it is not clear if this factor influences the effects of therapy, or the durability of these effects.

#### Treatment type

Casey and Berman (1985) compared the efficacy of two broad categories of child therapy, behavioral and non-

behavioral, and found that behavioral therapy produced larger effects. However, this difference was shown to be related to outcome measures particular to behavioral treatments. Casey and Berman (1985) reported that several behavioral treatment studies employed outcome measures that were inappropriate due to their similarity to the activities used in treatment. For example, for a treatment which reinforced good performance on a matching-to-sample task, a matching-to-sample task as a measure of outcome was deemed inappropriate. When studies of this nature were excluded ( $N = 29$ ), the behavioral and nonbehavioral treatments were of equal efficacy. Weisz et al. (1987) reexamined the studies excluded by Casey and Berman (1985). They reported that, while they were in agreement on many excluded cases, there were others wherein outcome measures similar to treatment activities were fair and necessary measures of treatment efficacy. The number of studies Weisz et al. (1987) judged to have been unfairly excluded by Casey and Berman (1985) was not specified. Nonetheless, like Casey and Berman (1985), Weisz et al. (1987) reported no significant difference in effects among the 21 distinct type of treatments they specified when this adjustment was made. However, it is not clear whether treatment type impacts the FU effects of child psychotherapy.

#### Mode of treatment delivery

Past reviewers have noted the lack of information

regarding the influence of individual versus group treatment delivery on treatment effects (Barnett, Docherty, & Frommelt, 1991; Kazdin, Bass, Ayers, & Rodgers, 1990). Thus, like problem severity, mode of treatment delivery is viewed by many psychotherapy researchers as a potential source of variance in treatment effects (Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz et al., 1992). Wiesz et al. (1987), reported larger PT effects for treatments administered individually rather than in groups, however the difference was not statistically significant. The impact of this factor on the FU effects of child psychotherapy has not been assessed in a cumulative fashion.

#### Number of treatment sessions

The "dosage of treatment" differs across outcome studies, and the potential impact of this variable is obvious. Often, reviewers assess the impact of treatment duration measured in weeks (e.g. Casey & Berman, 1985; Weisz et al., 1987). However, the real issue is a matter of the influence of the amount of treatment, of which number of treatment sessions is the more appropriate measure. Again, amount of treatment has been pointed to as a source of variability in treatment outcome (Kazdin, Siegel, and Bass, 1990; Weisz et al., 1992). Casey and Berman (1985) reported that length of treatment was negatively related to the PT effects of child interventions, but that this finding might be explained by the tendency of shorter studies to employ

outcome measures which produce larger effect sizes. The impact of the number of treatment sessions on the FU effects of child psychotherapy has not been previously addressed, however.

### Therapist experience

The influence of therapist experience on psychotherapy efficacy has been extensively studied. Durlak (1979) reported that professional therapists (Ph.D.'s, M.A.'s) are no more effective than professionally supervised paraprofessionals (teachers, parents) in administering psychotherapy to adults. Nietzel and Stuart (1981), drew similar conclusions from a re-examination of the sample of studies cited by the Durlak (1979) review. Casey and Berman (1985) found no effect for therapist experience on the efficacy of child psychotherapy, while Weisz et al. (1987) reported an interaction effect for therapist experience across some therapy types. As many studies employ trained teachers, parents, or other paraprofessionals to administer treatments, it is important to examine the influence of this factor on the long-term effects of child psychotherapy.

### Type of outcome measure

The influence of type of outcome measure on treatment effect size estimates has been noted by many reviewers (e.g. Smith et al., 1980; Casey & Berman, 1985; Weisz et al., 1987). Outcome measures vary across many factors, such as construct measured, sensitivity to change in the relevant



construct, and reliability of measurement. Thus, variance in treatment effects may be a result of these differences. In such cases, this variance is an artifact as far as clinicians are concerned, and should be noted as such.

Casey and Berman (1985) concluded that outcome measures unique to behavioral treatments had such an impact on PT effects in children as to produce the appearance of larger effects for behavioral compared to nonbehavioral treatments. Wiesz et al.(1987) also acknowledged the significant impact of outcome measures on PT effects, though not to the degree reported by Casey and Berman (1985). Thus, the potential for variability in FU effects in children due to different types of outcome measures has been convincingly demonstrated.

#### Clinical versus non-clinical studies

It has been argued that subjects from a true clinical inpatient or outpatient population are likely to differ from nonclinical populations across several factors, such as the presence of a formal diagnosis, severity of problems, and multiplicity of problems (Weisz et al., 1992). These proposed differences suggest that the two groups represent distinct populations, and thus the findings for one are not generalizable to the other. However, it also seems likely that many children in need of treatment never receive it, and thus many children with clinical levels of disturbance may be represented in large numbers within school popu-

lations. If this is the case, the use of screening instruments often employed in research to identify appropriate candidates for treatment would likely result in samples being drawn that closely resemble clinical populations.

Previous meta-analyses of the efficacy of child psychotherapy reported no reliable differences in PT effects between clinical and non-clinical samples (Casey and Berman, 1985; Weisz et al., 1987). Whether these samples produce differing effects at FU, however, has not been reported.

Assessing the influence of the clinical versus non-clinical nature of study samples on the FU effects of child psychotherapy will not resolve the controversy surrounding possible differences between clinically and non-clinically based populations. Such an evaluation will, however, provide a basis for judgement of whether these differences, if they do exist, are a matter of concern for the generalizability of the non-clinical studies.

#### Published versus unpublished studies

As previously mentioned, samples of studies drawn exclusively from journal publications tend to overestimate treatment effects (Smith, 1980). It will be important to assess the degree to which, if any, published studies overestimate FU effects of child psychotherapy.

#### Quality of studies

As is the case in any integration of original research,

the findings of such an integration are only as good as the quality of studies within the sample (Kendall, & Norton-Ford, 1982). This can be thought of as a "garbage in - garbage out" phenomenon: if the studies upon which a meta-analysis is based are of poor quality, then the results of that meta-analysis cannot be trusted. Thus, it is important to gauge the quality of the research base upon which a meta-analysis is built, and to determine how this factor influences the results.

### Hypotheses

The effects of child psychotherapy are expected to, on average remain stable across the FU interval. Thus, PT ES is predicted to be the strongest predictor of FU ES.

Although the influence of all of the previously mentioned variables will be assessed, two previous meta-analyses of the PT effects of child psychotherapy (Durlak, Lampman, & Wells, 1990; Durlak, Fuhrman, & Lampman, 1991) provide an empirical foundation for predictions regarding the primary moderators of FU effects. Durlak et al. (1990) reported that, as predicted, the primary moderators of the PT effects of child psychotherapy were the type of outcome measure used and the general type of treatment implemented.

Durlak et al. (1991) reasoned that the effects of cognitive-behavior therapy for children, a treatment that consists of facilitating the development and use of cognitive strategies to guide behavior, would be moderated by

the degree of advancement of a child's cognitive skills. Thus, as children in higher stages of development have higher cognitive skills, a child's developmental stage should be the greatest moderator of outcome for cognitive-behavior therapy. As previously mentioned, Durlak et al. (1991) used children's ages as an estimate of their developmental levels in this meta-analysis. They reported that the results were in line with this hypothesis, in that estimated stage of development was the greatest moderator of treatment effects in cognitive-behavior therapy for children.

In the current study, the effects of child psychotherapy are expected to be sustained through FU assessment, and to continue to be influenced by these primary moderating variables. Thus, treatment type is expected to be the primary moderator of the FU effects. Furthermore, the type of outcome measure used is expected to moderate the effects of all types of treatment as reported by Durlak, et. al (1990). It is possible that children's developmental levels may moderate outcome, and this possibility will be investigated. However, this variable will not be included in the primary model, as the type of outcome measure used has been shown to moderate outcome across more treatment modalities than has children's developmental levels. To continue, FU effects are expected to remain homogeneous (as measured by the Q statistic) within these subgroups, and are expected to

demonstrate no statistically significant differences from those at PT.

## CHAPTER III

### METHOD

#### Studies Reviewed

Studies relevant to the meta-analysis were reports appearing through the end of 1991 in which some form of psychotherapy for maladapting children (ages  $\leq 13$ ) was compared with a control group both at PT and after a FU interval. The following types of studies were excluded: those assessing only academic outcomes, and those assessing drug therapy or family therapy.

Psychotherapy was defined as any planned intervention designed to decrease maladaptive behavior, distress, or psychological symptoms or to improve adaptive functioning or prosocial behavior. Furthermore, the treatment had to be directed at children manifesting some degree of behavioral or social maladjustment, however defined (Durlak et al., 1990). Thus, the children who received treatment varied in both the type and severity of their dysfunction.

The search procedure consisted of three parts. First, previously obtained studies and reviews were scrutinized. Second, each article in 15 journals most likely to contain child psychotherapy research was examined (see Appendix A). Third, references of each included study were inspected.

While the search for published studies was not categorically exhaustive, it is likely that the sample represents a decided majority of relevant publications.

The potential for a publication bias within the child psychotherapy literature required that a representative sample of unpublished doctoral dissertations be obtained as well. To ensure that a representative sample was included, the number of relevant dissertations within the review period had to be determined.

Computer and manual searches of Dissertation Abstracts yielded approximately 660 citations between 1960 to 1991. Citations were divided into four year periods, and a 10% random sample of dissertations, stratified according to the year of completion, was targeted for review. Of this 10% random sample, 8 comparisons were relevant to the present review.

To summarize, search procedures produced a total of 107 comparisons meeting the inclusionary criteria. Of these comparisons, 101 were from journal or book publications, and 8 were from unpublished dissertations (see Appendix D).

### Procedure

#### Coding of studies

Studies were coded on 46 variables. Coded variables fell under one of the following general categories (see Appendix B): (a) study characteristics (e.g., publication status of study, number of comparisons included); (b) design

characteristics (e.g., type of design, assignment to groups, types of controls); (c) subject population characteristics (e.g., presenting problem, severity of problem); (d) therapist characteristics (e.g., training level); (e) treatment characteristics (e.g., general treatment approach, length of treatment); (f) outcome measure characteristics (e.g., rater perspective, dimension of adjustment assessed); or (g) characteristics of effect size calculations (e.g., statistics reported in each study).

While most of the coded variables are self-explanatory, those which, according to previous research (e.g., Casey & Berman, 1985; Durlak et al., 1991; Weisz et al., 1987), might moderate treatment effects require an explanation. First, participant's presenting problems were coded in two ways: first, problems were classified as one of 12 different categories along a continuum reflecting internalizing, externalizing, or mixed symptomatology, and then collapsed into one of these three general types of psychopathology to ensure adequate cell sizes (Achenbach & Edelbrock, 1983); and second, the severity of problems was coded as mild, moderate, or unknown. The internalizing categories included social isolation, fears and phobias, anxiety, enuresis, somatic problems, depression, or some combination of these. Externalizing categories consisted of impulsivity or hyperactivity, general behavior problems, inadequate social skills, or a combination of these. And



finally, mixed symptomology included any combination of internalizing and externalizing categories, and any combination of academic difficulties with either and internalizing or externalizing category. As result, coding procedures identified treated children in terms of both the severity and type of their maladjustment.

Problem severity was coded based on the diagnostic criteria and procedures used in each study. Severe psychopathologies, such as psychosis and autism, were not represented in the current sample because no investigations of these problems were identified by the search procedures. Problems were coded as moderate in severity when scores on a commonly used normed measure fell within the clinical range (e.g., error rate  $\geq 7$  and mean latency  $< 8.5$  seconds on the MFF, Kendall & Finch, 1977), when they were independently diagnosed by a pediatrician, psychiatrist, or physician (e.g., McGillivray, Cummins, & Prior, 1988), or a combination of these two occurrences (e.g., Hampe, Noble, Miller, & Barrett, 1973). Problems which were described in such a way as to suggest minimal distress or disturbance, such as a fear of snakes (Weissbrod & Bryan, 1973) or low rates of interacting with peers as rated by preschool directors (Keller & Carlson, 1974) were coded as mild in severity. Finally, studies which did not provide adequate sample diagnosis information to be classified as either moderate or mild in problem severity were coded instead as unknown

problem severity.

The developmental level of the children was estimated based on age such that preoperational corresponded to ages less than seven years ( $< 7$ ), concrete operational to ages seven to less than eleven years ( $7 - < 11$ ), and formal operational to ages eleven through thirteen years (11 to 13).

Treatment type was coded in two ways. First, to obtain the highest degree of specificity as possible, each study was coded as utilizing one of nineteen therapeutic techniques, such as reinforcement, desensitization, modelling, sociodrama, nondirective (Rogerian) techniques, and psychodynamic principles. Next, to ensure adequate cell sizes for analyses, these treatment types were collapsed into one of three general categories: behavioral, cognitive behavioral, or non-behavioral. In addition, the mode of treatment delivery was coded as individual, group, mixed, or unknown. Thus, coding for treatment was done first with as much specificity as possible, but was then combined to create broader categories with large enough cell sizes to allow for meaningful analyses.

The experience of the therapists utilized in each study was coded as mental health professionals (Ph.D. in psychology, M.S.W., M.D. in psychiatry, or M.A. in school guidance), professional trainees (graduate students in psychology, interns, practicum students, or psychiatric

res-idents), paraprofessionals (parents, teachers, other non-professionals), mixed, or unknown.

Outcome measures were coded several different ways as well. First, an outcome measure was coded into one of six general descriptive categories (e.g., normed measure, sociometric; see Appendix B). Second, they were coded regarding whether the measure assessed a specific or general impact of treatment, and which dimension(s) of adjustment was being evaluated (e.g., overt behavior, personality traits such as anxiety or locus of control, or cognitive processes such as cognitive problem-solving skills or cognitive tempo).

The context of each study was coded as either clinical or non-clinical. Clinical studies included only those utilizing either clinical inpatients or clinical outpatients independently seeking treatment. Non-clinical studies included samples of project volunteers, or those chosen through some problem-oriented screening process, those of convenience, and those of mixed characteristics. Finally, samples whose sources were not described were not included under either classification.

Finally, features relevant to the experimental quality of each study were coded, such as experimental design, type of control group, and the attrition rate from pretreatment to PT.

### The meta-analytic technique

The meta-analytic technique involves deriving standardized scores reflecting treatment effectiveness from original studies. This is accomplished by calculating the PT or FU difference between treatment and control group means for each study and then dividing this sum by the pooled standard deviation of the two groups. The resultant statistic, called an effect size, provides a quantitative description of both the size and direction of treatment effects for each study. These effect sizes, as they are expressed in standardized units, can be averaged and compared across studies. Furthermore, by coding for variables which vary across studies, the moderating effects of these variables can be assessed. Thus, the meta-analytic method provides a means for resolving inconsistent findings by (a) locating the central tendency of the effects of all related treatment efficacy studies and (b) determining if the inconsistencies are due to the moderating effects of some variable or variables on which the studies differ. Several of the advantages meta-analysis provides relative to alternative review techniques are well defined by Bryant (1986):

". . . the traditional qualitative review is largely subjective and provides little or no statistical information about the strength of observed effects. Furthermore, other methods of quantitative review, such as a simple 'vote count' that categorizes studies' outcomes as positive, negative, or zero effects, can produce misleading 'no difference' conclusions, or Type II errors, because of low statistical power (Hedges & Olkin, 1980; Light & Pillemer, 1984; Light &

Smith, 1971). Research synthesis allows more systematic investigation of the mean and variance of effect sizes. Thus, the main strength of research synthesis is that it provides a quantitative index of treatment effects expressed in a metric that is comparable across studies." (pp. 2-3).

By aggregating effect sizes in this manner, inconsistencies among findings of original studies can be resolved by deriving a grand mean effect size indicating the central tendency of the included studies. For example, when the number of available studies reporting significant effects is equal to that reporting insignificant effects, the conventional "vote counting" review method would suggest that the treatment being studied produces inconsistent results. However, with the use of meta-analysis, the central tendency of effects among these studies can be expressed quantitatively.

Despite the advantages in research integration afforded by meta-analysis, it is not without its limitations. As previously mentioned, the technique relies upon an adequate base of quality original research in order to produce results that are of any use. Furthermore, as meta-analysis uses treatment comparisons as the unit of analysis, information about the performance of specific individuals within each study is obscured. Thus, only information relating to group averages can be obtained.

In addition to these limitations, the calculation of a grand mean effect size among studies with different subject

and treatment characteristics can be misleading, as systematic variation within these subgroups can occur without being reflected by a change in the mean. When studies from different underlying populations are lumped together in this way, one is left with a comparison of "apples and oranges." To avoid this problem, studies should be aggregated in homogeneous groups. A group of outcome effects is judged to be homogenous when variability in effects obtained by these studies is due to sampling error rather than systematic differences among the studies in addition to sampling error (e.g., they are derived from the same underlying population). Statistical procedures have been developed to test for homogeneity among effect size estimates (Hedges & Olkin, 1985). In the absence of such a test, simply reporting the grand mean effect size could obscure real differences that might occur among individual effect size estimates. In other words, the true distribution could be multi-modal, in which case the grand mean is a misleading statistic.

#### Effect size calculations

Effect sizes for each comparison were calculated in one of twelve ways. In studies reporting means, standard deviations, and the number of subjects in each group, effect sizes were calculated using the following formula:

$$\text{Effect size} = \frac{\text{Mean (treatment)} - \text{Mean (control)}}{\text{pooled standard deviation}},$$

wherein the pooled standard deviation is the number in the

treatment group minus one times the standard deviation of the treatment group, plus the number in the control group minus one times the standard deviation of the control group, divided by the total number in both groups minus two.

For studies in which this information was not provided, procedures described by Wolff (1986) were used to calculate effect sizes (see Appendix C).

### Model testing procedures

To test for goodness of fit of the proposed model the categorical fixed effects approach developed by Hedges and Olkin (1985) was used. This procedure involves theorizing a model based on variables that are expected to moderate the effects. Comparisons are grouped according to moderating variables, and a Q statistic is calculated to assess the homogeneity of the effect sizes within these groupings. The Q statistic is calculated as a chi square ( $df = k - 1$ , where  $k$  is the number of studies included in the category).

Because homogeneity within each group is indicative of a good model (i.e. studies within each group are drawn from the same underlying population), a nonsignificant Q is the desired result. Homogeneity was tested at  $p \leq .05$ .

In addition to the test for homogeneity, a weighted regression can be used to identify variables which moderate treatment effects (Hedges & Olkin, 1985). The weighted regression is used in addition to the Q test for homogeneity because variables that do not produce homogeneity in ES

distributions may nonetheless moderate treatment effects (Shaddish, 1993). In this regression analysis, each ES is weighted by the study sample size. The analysis consists of entering each possible predictor into a weighted regression to identify the most significant predictor. Then, the most significant predictor is removed from the equation, and the process is carried out again to identify the next most significant predictor. This process is continued until the change in sum of squares regression fails to exceed  $Q_{critical}$  at  $p \leq .01$  ( $df = k-1$ , where  $k$  = the number of predictors in the step), at which point the predictors are no longer considered significant. Model specification is reached only when sum of squares residual fails to exceed  $Q_{critical}$  at  $p \leq .01$  ( $df = N-k$ , where  $k$  = the number of predictors included in the step, and  $N$  = the total number of studies included in the analysis).

### Weighting procedures

As studies with larger sample sizes contain less sample variance, they produce more accurate effect size estimates (Hedges & Olkin, 1985). Thus, it follows that studies with larger sample sizes should be given greater weights when pooling effect sizes. This will result in a less biased estimator than is obtained with unweighted effect sizes. Because of these considerations, effect sizes in this review were weighted using the calculations recommended by Hedges and Olkin (1985) when using the  $Q$ -test for homogeneity and



the weighted regression equation.

### Outlier analysis

Because outlier effect sizes might prevent the distributions under investigation from reaching homogeneity, it is essential to identify any such occurrences and to eliminate them from subsequent analyses (Hedges & Olkin, 1985). Thus, an outlier analysis was incorporated into the homogeneity testing process. In line with Hedges and Olkin (1985), any ES which disproportionately contributes to the Q statistic was considered an outlier. Determining whether an ES contributes disproportionately to the Q statistic requires that this value be examined, and a judgement be made regarding whether or not it is disproportionate to that of other ES's.

Because both ES and sample size influence each comparison's contribution to the Q statistic, a scatterplot of ES by sample size was constructed in order to identify comparisons which might meet the outlier criterion. Outliers were identified at PT rather than at FU in order to facilitate the establishment of an initial group of homogeneous studies that could be followed over time. The application of this procedure will be taken up later in this text.

## CHAPTER IV

### EXPERIMENT RESULTS

#### Characteristics of Reviewed Studies

Table 1 contains descriptive characteristics of the studies under review (after the removal of outliers as described below) in terms of client, treatment, and methodological variables. As can be seen, the average study involved a sample approximately 9.1 years of age, consisting primarily of white males (65.1%) with moderately severe pathology (48.5%), the most common of which were externalizing difficulties (45.5%). Most of these children were treated using either a behavioral (35.6%) or cognitive-behavioral (42.6%) approach, and were seen with equal frequency in both individual and group contexts for an average of 12.62 sessions. While therapists were of various levels of training, professionals (36.6%) were more common than paraprofessionals (17.8%) or professional trainees (11.9%).

Most of the included studies utilized random assignment to experimental conditions (82.3%), and had an attrition rate of less than 10% (81.6%). In addition, 41.6% of the studies included a placebo control group, 19.5% used a normed outcome measure, 89.4% used multiple outcome

Table 1

Selected Characteristics of Studies (N=101)

Characteristic	Proportion or Mean
Age (mean/sd)	9.1 (2.54)
Males	65.1%
Minorities Represented	15.8%
Problem Type	
Internalizing	28.7%
Externalizing	45.5%
Mixed	25.7%
Problem Severity	
Mild	21.8%
Moderate	48.5%
Uncertain	28.7%
Sample Population	
Clinical, Moderate Severity	13.3%
Non-clinical, Moderate Severity	34.7%
Non-clinical, Mild Severity	22.4%
Non-clinical, Unknown Severity	29.6%
Academic Problems Present	17.8%
Number of Treatment Sessions (mean/sd)	12.62 (13.71)
Average Length of FU in Weeks (mean/sd)	26.68 (41.91)

Table 1 (continued)

Selected Characteristics of Studies (N=101)

Characteristic	Proportion or Mean
<b>Treatment Type</b>	
Behavioral	35.6%
Cognitive-Behavioral	36.6%
Non-Behavioral	27.7%
<b>Mode of Delivery</b>	
Individual	44.6%
Group	43.6%
Mixed	11.9%
<b>Experience Level of Therapists</b>	
Professionals	36.6%
Graduate Students	11.9%
Paraprofessionals	17.8%
Mixed	14.9%
Unknown	18.8%
<b>Dimension of Adjustment Assessed</b>	
Behavioral	45.8%
Personality	18.1%
Academic/Cognitive Skills	15.3%
Sociometrics	11.1%
Other	9.7%

Table 1 (continued)

Selected Characteristics of Studies (N=101)

Characteristic	Proportion or Mean
Utilized Random Assignment	82.3%
Attrition Rate Less Than Ten Percent	81.6%
Utilized a Placebo Control Group	41.6%
Utilized a Normed Outcome Measure	19.5%
Utilized Multiple Outcome Measures	89.4%
Assessed Generalized Impact of Treatment	39.9%

measures, and 39.9% assessed the generalized impact of the treatment.

Due to the lack of specificity in the problem categories (i.e., externalizing, internalizing, and mixed) and the treatment categories (i.e., behavioral, cognitive-behavioral, and non-behavioral), a more detailed description of the contents of these categories would be beneficial. To begin with, the "externalizing" category referred to the following types of difficulties: impulsivity or hyperactivity; general behavior/management problems such as noncompliance, temper tantrums, fighting with siblings, negativity, complaining, and crying; a combination of hyperactivity and general behavior/management problems; and inadequate social skills. Of these problem types, the most common were impulsivity/hyperactivity (43.5%) and general behavior/management problems (30.4%).

Internalizing disorders included social isolation, fears or phobias, anxiety, somatic problems such as abdominal pain, depression, or some combination of these. While the most frequent problem classified as "internalizing" was social isolation (34.5%), fears and phobias (17.2%) and somatic problems (17.2%) also constituted a substantial portion of these cases.

Finally, the "mixed" category contained samples exhibiting some combination of internalizing, externalizing, and/or academic difficulties. Examples from this category

include samples described as socially maladjusted with reading deficiency, conduct disordered with academic difficulties, behaviorally noncompliant with separation anxiety, and aggressive with verbal comprehension difficulties. As can be seen, this category contained a heterogeneous mix of problem combinations. Thus, no one specific combination of problems constituted a substantial portion of this category.

With regard to the treatment categories, behavioral treatments included reward oriented reinforcement, a combination of reward and punishment, systematic desensitization, modeling, relaxation and biofeedback, and reinforcement combined with modeling. Modeling (30.3%), reinforcement combined with modeling (22.2%), and reward combined with punishment (13.9%) constituted the majority of these studies. Non-behavioral treatments encompassed miscellaneous non-directive interventions (4.0%), verbal interaction with exercise (5.9%), activity therapy (17.9%), Rogerian therapy (17.9%) and psychodynamic therapies (14.3%). Finally, the cognitive-behavioral treatments included interventions designed to modify cognitive processes in an effort to change behavior. Various self-instructional procedures figured prominently in these studies.

#### Posttreatment Versus Follow-up Comparisons

To begin with, a grand weighted mean ES and a 99% confidence interval (CI) were calculated for both PT and FU

assessments. Note that the mean ES's and CI's at PT (N=109, ES= 0.21, CI= 0.16 - 0.27) and at FU (N=109, ES= 0.23, CI= 0.18 - 0.28) are similar, suggesting that treatment effects remained stable across the FU interval. However, looking within the PT and FU assessments, results of the Q-test indicated that the distributions of effect sizes were not homogeneous at either PT ( $Q = 251.81, p < .01$ ) or FU ( $Q = 235.20, p < .01$ ). This finding raised three possibilities: (1) the presence of outlier ES's within the distributions may have affected the homogeneity of variance; (2) the presence of moderator variables prevented homogeneity; and (3) a combination of the first two possibilities. Thus, no conclusions could be drawn until all of these possibilities were evaluated.

#### Outlier Analysis

The procedures previously described were used to identify any outliers which might have contributed to the heterogeneity of ES's at PT and FU. This process identified eight studies. Four of these were investigations of non-behavioral treatments of children with a mix of internalizing and externalizing difficulties. Another was of non-behavioral treatment of impulsivity, and the final three studies were of behavioral treatments for children who were either social isolates, low in academic achievement, or impulsive. Each of these was omitted from subsequent analyses.



### Re-analysis of PT and FU Comparisons

Having discarded the outliers, PT and FU effect sizes were again compared across the remaining studies. Once again, no difference was found between the grand mean ES's at PT ( $N=101$ ,  $ES= 0.44$ ,  $CI= 0.36 - 0.53$ ) and FU ( $N=101$ ,  $ES= 0.45$ ,  $CI= 0.36 - 0.54$ ). However, as occurred in the previous comparison, neither the PT distribution ( $Q= 152.8$ ,  $p < .01$ ) nor the FU distribution ( $Q= 146.6$ ,  $p < .01$ ) reached homogeneity. This suggests that ES's at both PT and FU are moderated by one or more unidentified variables. As a result, it is inappropriate to make a comparison between these two ES distributions, and any conclusions based on this comparative analysis would be premature.

### Analyses of Moderator Variables

It was expected that ES's at PT and FU would fail to reach homogeneity, as previous research and reviews have indicated that effect sizes could be moderated by any one or combination of several variables (Barnett, Docherty, & Frommelt, 1991; Casey & Berman, 1985; Durlak et al., 1991; Kazdin, 1993; Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz et al., 1987; Weisz, Weiss, & Donenberg, 1992). Moderator variables in meta-analysis are those across which ES's vary systematically. When broken across moderator variables, ES's should be distributed homogeneously (Hedges & Olkin, 1985).

The next step in the investigation was to identify

which variables moderate PT and FU ES's, and to compare the grand mean PT and FU ES's across these variables. Treatment type and presenting problem, the two variables constituting the most salient differences among the studies under review, were the primary variables expected to moderate FU ES's. However, all of the variables included in Table 1, based upon the above citations, could be considered potential moderator variables, and each was eventually included in this analysis.

Table 2 contains the mean ES, CI, and  $Q_{\text{observed}}$  for each level of treatment type and presenting problem at PT and FU assessments. As can be seen, treatment type was broken into behavioral ( $n=36$ , PT  $Q=44.04$ , FU  $Q=39.57$ ), cognitive-behavioral ( $n=43$ , PT  $Q=58.10$ , FU  $Q=62.38$ ), or nonbehavioral studies ( $n=22$ , PT  $Q=32.44$ , FU  $Q=31.30$ ), while presenting problem was divided into internalizing ( $n=29$ , PT  $Q=26.05$ , FU  $Q=42.18$ ), externalizing ( $n=46$ , PT  $Q=67.65$ , FU  $Q=55.14$ ), or mixed ( $n=26$ , PT  $Q=35.37$ , FU  $Q=34.21$ ). The  $Q$  statistics testing the homogeneity of ES's at PT and FU for each of these comparisons failed to reach significance, indicating that ES's were homogeneously distributed across these variables. This result suggests that both PT and FU ES's are moderated by the type of treatment and the type of problem being treated.

With these two moderator variables identified, comparisons between PT and FU ES's were finally facilitated.

Table 2

Homogeneity Test of Moderator Variables

Treatment type	$Q_{obs}^*$	weighted ES	99% CI
Behavioral ( $\underline{n}=36$ )			
posttreatment	44.04	.50	.35-.64
follow-up	39.57	.54	.39-.68
Cognitive-behavioral ( $\underline{n}=43$ )			
posttreatment	58.10	.57	.42-.72
follow-up	62.38	.52	.37-.67
Non-behavioral ( $\underline{n}=22$ )			
posttreatment	32.44	.21	.05-.38
follow-up	31.30	.25	.09-.42
Problem type			
Internalizing ( $\underline{n}=29$ )			
posttreatment	26.05	.61	.42-.80
follow-up	42.18	.66	.47-.85
Externalizing ( $\underline{n}=46$ )			
posttreatment	67.65	.54	.41-.67
follow-up	55.14	.47	.34-.60
Mixed ( $\underline{n}=26$ )			
posttreatment	35.37	.22	.07-.37
follow-up	34.21	.30	.15-.45

\*Note. Q was tested at  $p < .01$ , and failed to reach significance in each of these comparisons.

CI= confidence interval.

ES= effect size.

As can be seen in Table 2, ES's are relatively stable from PT to FU across each level of both treatment and problem types, and the 99% confidence intervals within each level of both of these variables at both PT and FU are similar. These results suggest that treatment effects realized at PT are stable across the FU interval, and equally so across each level of both moderator variables.

#### Modeling FU ES's

By identifying the moderator variables (presenting problem and treatment type), we have specified variables across which the effects of treatment are homogeneously distributed. While it stands to reason that these variables are predictive of FU ES's, it is also possible that other variables may also predict FU ES's (Shadish, 1994). In addition, continuous variables which are treated categorically for use in the Q test, such as age, may lose much of their predictive power in the process, as categorization of these variables restricts their range of variability. As a result, other variables which are not identified as moderators by the homogeneity tests in Table 2 may still account for a significant degree of variance in FU ES's.

In light of these considerations, the next analyses involved using a weighted regression to build a model to account for the variance in FU ES's. This procedure has been described above: all variables from Table 1 were entered into the regression individually to determine the

strongest predictors of FU ES's. The strongest predictor at each step in the regression was determined, and then the next strongest predictor was determined by entering the remaining variables into the regression one at a time. This process was continued until the change in sum of square regression ceased to reach significance as measured by  $Q$  at  $p < .01$ .

All variables included in Table 1 were entered into the weighted regression. In addition, PT ES was entered into the equation as well. This was done because if, as expected, treatment effects are stable over time, PT ES should be a significant predictor of FU ES. Table 3 presents variables which were significant predictors of FU ES. As expected, PT ES was the most significant predictor ( $R = .72$ ), accounting for 52% of the variance.

Two additional significant predictors of FU ES emerged. These were problem type ( $R = .76$ ,  $R^2_{\text{change}} = .05$ ), and total number of therapy sessions ( $R = .77$ ,  $R^2_{\text{change}} = .03$ ). It is noteworthy that PT ES accounts for a majority of the variance in FU ES (52%), while problem type and number of sessions account for relatively little (5% and 3%, respectively). The other variables assessed were not significant predictors. These were average age, percentage of males, ethnicity, problem type, problem severity, source of participants (i.e., clinical or non-clinical), presence of academic problems, length of FU interval, general type of

Table 3

Weighted Regression Predicting FU ES's

Step	Var	Mult R	R <sup>2</sup>	R <sup>2</sup> change	SSregress	SSresid
1	PTES	.72	.52	.52	534.57*	483.89*
2	PROBTP	.76	.57	.05	583.32*	435.14*
3	NSESS	.77	.59	.02	604.34*	414.12*

\* significant at .01 level. Significant SSregress is desired to identify predictors of ES, but nonsignificant SSresid is desired to indicate model specification. Var= variable, Mult R= multiple R, SSregress= sum of squares regression, SSresid= sum of squares residual, PTES= posttreatment effect size, PROBTP= general type of problem, NSESS= number of therapy sessions

treatment, mode of delivery (i.e., individual or group), experience level of therapist, dimension of adjustment assessed (i.e., personality, behavioral, academic/cognitive skills, sociometrics, or other), and the methodological quality of the included studies (i.e., use of random assignment, use of a true placebo control group, attrition rate less than 10%, use of multiple outcome measures, use of a normed outcome measure, and assessment of the generalized impact of the treatment under study). These data suggest that PT ES is the most important indicator of FU effect, although two other variables do contribute to a much lesser degree. This lends further support to the notion that treatment effects realized at PT, on average, remain stable across FU intervals.

#### Examining Change from PT to FU

Although FU ES's overall do not change over time, there is variability among the ES's over time. Some of the included studies reported a decline, others reported an increase, and still other reported no change in ES from PT to FU. For example, as previously mentioned, three FU studies of the treatment of social relation problems in children produced different results: Gottman (1977) found that therapeutic effects dissipated over time, Jakibchuck and Smeriglio (1976) that the effects increase, and Weinrott, Corson, and Wilchesky (1979) reported that the effects remain stable after treatment termination.

In order to describe the current sample in terms of the degree of change from occurring from PT to FU, a change score was calculated for each study by subtracting the FU ES from the PT ES. Thus, a positive change score indicated that the ES declined from PT to FU, and a negative change score indicated that the ES increased from PT to FU. The mean change score for all studies was .01 (sd= .38) indicating that, on average, ES's decreased one hundredth of a standard unit from PT to FU. Change scores ranged from -1.05 to 1.25.

Repeated measures ANOVA's or regression analyses (depending on the variable) were used in order to determine if changes in ES's from PT to FU could be explained by the variables listed in table 1. As a total of 19 variables were evaluated in these analyses, a Bonferroni correction was used to prevent obtaining chance findings, setting the  $p$  level at .002. In analyses involving a repeated measures ANOVA, a significant interaction between the variable under consideration and time indicated that that variable influenced change in ES. In analyses involving a regression, significant predictors of FU ES after covarying out the effect of PT ES indicated significant predictors of change in ES.

All of these analyses were non-significant, suggesting that change in ES from PT to FU could not be accounted for by the variables listed in table 1.



### Length of FU Interval

It seems intuitive that a FU assessment performed after one interval of time is not necessarily equivalent to a similar assessment performed after a different interval. For example, it would be difficult to contend that an assessment taken one week after treatment termination bears the same meaning as an assessment taken two years after treatment termination, as the two year assessment would appear to provide a much more rigorous test of the durability of treatment effects than the one week assessment. Thus, a closer look at the range of FU intervals represented in the current investigation is called for.

As noted Table 1, the mean length of FU intervals as measured in weeks was 26.7 ( $SD = 41.91$ ). The length of the FU interval varied considerably among studies ranging from 1 week ( $n = 3$ ) to 260 weeks ( $n = 1$ ). The modal interval length was 4 weeks ( $n = 13$ ), closely followed by 52 weeks ( $n = 12$ ). The median interval length was 10 weeks ( $n = 6$ ), indicating that more than half of the studies contained FU intervals of 2.5 months or less.

Despite the finding that, as has been previously noted, length of FU was not a significant predictor of FU ES, studies were split into two groups: those with a FU interval less than one year ( $n = 79$ ) and those with a FU interval equal to or greater than one year ( $n = 22$ ). These two groups were compared across each of the variables listed

in Table 1 using Chi squares and oneway ANOVA's, but no significant differences were found. In addition, FU ES's of these two groups were compared using an ANOVA, and no significant differences were found. From these observations, it would appear that the length of the FU interval, either standing alone or in covariance with other variables included in this investigation, does not influence FU ES's.

#### Representativeness of the Study Sample

In order to determine the degree to which the studies used in the present meta-analysis are representative of child psychotherapy outcome studies in general, the study characteristics of the current sample ( $N = 101$ ) to a larger sample of studies which did not include FU across each variable listed in Table 1. This latter sample of studies ( $N = 416$ ) was drawn from that described in Durlak, Wells, Cotten, and Johnson (in press), and the selection procedures used were identical to those described in the current study with the exception of the inclusion of evaluations that did not report FU assessments.

Among all of these variables, the only significant difference was across treatment types ( $\chi^2 = \quad$ ,  $p < .001$ ), indicating that studies with FU assessments differed from the larger body of child psychotherapy outcome studies by including more assessments of cognitive-behavioral techniques (35.6% compared to 15.3%) and fewer assessments of nonbehavioral techniques (28.8% compared to 46.3%). No

other differences were noted between these two groups of studies. This suggests that the literature examining FU effects is similar to the body of child psychotherapy outcome literature as a whole with the exception of the proportion of general treatment types represented in each.

## CHAPTER V

### DISCUSSION

#### Summary

The current study sought to paint an integrated picture of the effects of child psychotherapy over time as revealed by the current treatment outcome literature. In the development of this picture, several questions arose: Is the FU outcome literature homogeneous with regard to the distribution of reported ES's (after eliminating outliers)? If not, can variables that moderate FU ES's be identified? Are ES's, on average, stable from PT to FU when broken across these moderator variables? What variables are the strongest predictors of FU ES? Do significant numbers of studies report a change in ES from PT to FU? If so, can this change be accounted for? And finally, is the FU literature notably different from the outcome literature as a whole?

FU ES's were found to be heterogeneously distributed across all FU studies. This finding suggested that, as expected, FU ES's were moderated by some other variable or variables. In other words, there appeared to be variables that influence FU ES's to the extent that FU ES's were homogeneously distributed across these variables only.

The second question logically follows from the first. If FU ES's are influenced by other variables as described above, what are these variables? Are they clinical variables, such as problem type or developmental level of the child? Or instead, are they research methodological variables, such as attrition rate or type of control group utilized? If the moderator variables are of the former type, the implication is that FU ES's are influenced in a theoretically, or at least clinically relevant manner. This interpretation has been made regarding other areas of the child psychotherapy outcome literature. For example, Durlak et al. (1991) suggested that PT ES's of cognitive behavioral treatments were moderated by the recipients' developmental levels. In addition, Durlak et al. (1990) presented evidence that PT ES's for child treatments in general were moderated by the dimensions of adjustment being assessed.

If the moderators are of the latter type, then it would appear as if FU ES's were being influenced by variables with no inherent clinical or theoretical value, to the extent that differences in FU ES's across meaningful variables were washed out. Obviously, such a conclusion would bring the validity of the entire literature into question. Some previous reviewers have suggested that the literature is in fact questionable for this very reason (e.g., Kazdin et al., 1990; Weisz et al., 1992). Fortunately, by coding studies across both types of variables, I was able to address this

issue.

The results suggest that FU ES's are moderated by general type of treatment, and general type of problem being treated. No other variables emerged as moderators in this analysis, although several variables of potential importance were excluded because they were continuous (e.g., length of FU interval, PT ES, and number of treatment sessions). Despite the fact that the hypothesis regarding developmental level as a moderator variable was not supported by these results, they do indicate several noteworthy things. First, theoretically and clinically relevant variables were more influential than "nuisance" variables in determining FU ES's. Thus, the literature would appear to be of sufficient methodological rigor to produce meaningful results.

Second, it appears that there are differences among the effects of the multitude of possible treatment approaches, and that very rough treatment descriptions, such as behavioral, cognitive-behavioral, and non-behavioral provide sufficient categories for capturing these differences. Such a finding suggests that each of these categories contain approaches that differ in their "active ingredients" or "active processes" from each of the other categories, an idea that is currently embroiled in considerable controversy (e.g., Orlinsky & Howard, 1986; Shapiro, Harper, Startup, Reynolds, Bird, & Soukas, 1994; Stiles, Shapiro, & Harper, 1994).

Third, the results imply that presenting problems respond to treatment over time in a similar fashion to other problems from the same general category (i.e., internalizing, externalizing, or mixed), but in a distinct fashion from those in each of the other two categories. Again, it appears that differentiating types of problems even in the most general terms captures differences in their responsiveness to treatment.

As was previously discussed, identifying these moderator variables facilitated a meaningful comparison of PT and FU ES's. Each of the two moderator variables, treatment type and problem type, contained three levels: behavioral, cognitive-behavioral, and nonbehavioral treatment types, and internalizing, externalizing, and mixed problem types. Because both PT and FU ES's were homogeneously distributed across these variables, the average ES was the most unbiased estimator of the central tendency. Note in Table 2 that the greatest difference between average PT and FU ES's for any of these categories was a decrease of .07 from PT to FU for externalizing problems. Furthermore, note that the PT and FU 99% confidence intervals within each problem and treatment type are overlapping. These observations provide strong support for the notion that the effects of child psychotherapy, on average, remain stable over time.

To further test the idea that treatment effects are stable over time, a weighted regression analysis was used to

identify the strongest predictors. If treatment effects are indeed stable over time, the strongest predictor of FU ES should be PT ES. Thus, this was the expected finding. However, if another variable emerged as the strongest predictor of FU ES, conclusions regarding the stability of treatment effects over time would have to be qualified accordingly. In addition, the regression analysis provided a more powerful test of the influence of the continuous variables listed in Table 1 (i.e., number of treatment sessions, age of participants, and length of FU) that did not easily lend themselves to the categorization required by the Q test for homogeneity.

The results of the regression analysis further strengthened the argument for the stability of treatment effects over time, as PT ES emerged as the most significant predictor of FU ES. In addition, the finding that the number of treatment sessions is positively related to FU ES should prove to be of interest to both clinicians and policy makers. These data raise the issue that the number of treatment sessions necessary for immediate change may not be sufficient for lasting change. In other words, ten sessions may be sufficient for symptom amelioration from pretreatment to PT. In some cases, however, additional sessions may be necessary to ensure that the symptom does not return at some point after treatment termination.

Despite the many indications that child treatment



effects remain stable over time, several studies did report a change in average ES from PT to FU. The existence of these studies suggests that there are conditions under which treatment effects drop off significantly after treatment termination (mean change  $\underline{ES}$  =  $-.68$ ), and conditions under which treatment effects increase as time passes after termination (mean change  $\underline{ES}$  =  $+.64$ ). However, efforts failed to identify these conditions.

It is a significant finding that the conditions under which change in ES from PT to FU occurs could not be specified from the variables reported in Table 1. This suggests that perhaps variables other than the type reported here bear an influence on the FU effects of child psychotherapy. For example, it has been argued that the presence of family risk factors, such as poverty, a mentally ill parent, a substance dependent parent, divorce, and an authoritarian parenting style, contribute to or even cause psychopathology in children (Rutter, 1984; Rutter, 1985). Thus, it seems intuitive that the way a child is treated by his or her family, and the role the child occupies in his or her family, will influence his or her ability to sustain any benefits realized in psychotherapy. The current study suggests that the influence of family variables on the long term effectiveness of child psychotherapy is an area in need of study.

In addition, it has been suggested that the effects of

therapy are related to interpersonal processes occurring within therapy, and that we have failed to adequately describe these processes up to now (Shapiro et al., 1994). Thus, the study of process variables is an area of inquiry needing further development.

Finally, the child FU outcome literature generally appears to be similar to the child outcome literature as a whole in terms of the types of studies found in each. The one exception is that the FU literature contains a larger percentage of cognitive-behavioral interventions, while the literature as a whole contains a larger proportion of nonbehavioral interventions.

### Limitations

The findings of the current meta-analysis are subject to several limitations. These limitations will be discussed below.

First, because meta-analysis deals solely with average study scores rather than individual participants' scores, it is uncertain whether the trends evidenced in PT and FU ES's across studies would be replicated by the scores of individual participants in original research. Thus, the findings have limited relevance to an individual child in treatment. Second, many of the variables (e.g., study quality variables, problem severity, length of treatment, etc.) either required a notable degree of subjective judgement in their coding, or were reported infrequently and inconsistently.

Thus, the coding system, while as good as allowed by the current state of the literature, is often limited in terms of its specificity and its objectivity. Nonetheless, the coding of variables that seem to be promising in terms of predicting FU ES's needs to be refined and the variables studied further.

Finally, the child psychotherapy outcome literature as a whole has been criticized regarding its internal and external validity (i.e., Kazdin et al., 1990; Weisz et al., 1992). While many of the criticisms are disputable (see Durlak et al., in press), at least one holds true in the FU literature. First while there is a substantial number of FU studies in general, there is a limited number of studies regarding specific interventions (e.g., psychodynamic therapies, play therapy, skills training, desensitization, etc.) Obviously, the conclusions of the current meta-analysis can apply only to those intervention which are represented in the sample.

## Appendix A

The following journals were searched manually:

Journal of Clinical Child Psychology, Journal of Abnormal Child Psychology, Journal of School Psychology, Journal of Consulting and Clinical Psychology, Journal of Applied Behavior Analysis, Behavior Therapy, Behavior Modification, Cognitive Therapy and Research, American Journal of Community Psychology, Elementary School Guidance and Counseling, Psychology in the Schools, School Counselor, Journal of Counseling Psychology, Behavior Research and Therapy, and Journal of Community Psychology.

## Appendix B

## Coding Scheme for Included Studies

## I. Study Characteristics

- \_\_\_\_\_ 1. Study ID# (001-999)
- \_\_\_\_\_ 2. Year of publication
- \_\_\_\_\_ 3. Source of study
  - 1= published report
  - 2= unpublished report
- \_\_\_\_\_ 4. Total number of treatment groups
- \_\_\_\_\_ 5. Total number of comparisons
- \_\_\_\_\_ 6. Total number of outcome measures

## II. Design Features

- \_\_\_\_\_ 7. Type of design
  - 1= Pretest- Posttest with nonequivalent control group
  - 2= Posttest only with nonequivalent control group
  - 3= Randomized true experiment
  - 4= Other (e.g. matching)
- \_\_\_\_\_ 8. Group assignment procedures
  - 1= random
  - 2= matching
  - 3= available intact
  - 4= voluntary self-selection
  - 5= other
  - 6= not available
- \_\_\_\_\_ 9. Total sample size assigned  
(all treatment and control groups)
- \_\_\_\_\_ 10. Total sample size completed at posttest  
(all treatment and control groups)

## III. Subject Information

- \_\_\_\_\_ 11. Percentage of males in sample
- \_\_\_\_\_ 12. Mean age of subjects to the nearest tenth of a year

## Appendix B (continued)

- \_\_\_\_\_ 13. Ethnic characteristics
  - 1= majority or all white
  - 2= majority or all minority
  - 3= mixed
  - 4= unknown
  
- \_\_\_\_\_ 14. Special sample characteristics
  - 1= retarded
  - 2= learning disabled
  - 3= underachievers
  - 4= hospital/dental patients
  - 5= other
  - 6= unknown
  - 7= none
  
- \_\_\_\_\_ 15. Source of participants
  - 1= clinical inpatients
  - 2= clinical outpatients seeking treatment
  - 3= volunteers for special project
  - 4= subjects chosen through problem-oriented observation,  
measurement, or recommendation
  - 5= hospital/dental patients
  - 6= sample of convenience
  - 7= mixed/other
  - 8= unknown
  
- \_\_\_\_\_ 16. General seriousness of problem
  - 1= mild
  - 2= moderate to severe
  - 3= of uncertain nature/degree
  
- \_\_\_\_\_ 17. Target problem
  - 1= social isolation
  - 2= fears/phobias
  - 3= anxiety
  - 4= enuresis
  - 5= somatic problems
  - 6= depression
  - 7= other or mix of 1-6
  - (1-7 indicate internalizing symptomatology)
  - 8= impulsive/hyperactive
  - 9= non-compliant/management problem/behavior problem
  - 10= psychotic/autistic
  - 11= other or mix of 8-10
  - 12= social skills, undefined
  - (8-12 indicate externalizing symptomatology)
  - 13= mix of 1-12
  - 14= other/unknown

## Appendix B (continued)

## \_\_\_\_\_ 18. Academic learning problems

- 1= present
- 2= not present
- 3= unknown

## IV. Therapist Characteristics

## \_\_\_\_\_ 19. Number of therapists

## \_\_\_\_\_ 20. Experience level of therapists

- 1= mental health professionals (Ph.D. in psychology, social work; MD in psychiatry; school guidance counselor)
- 2= professional trainees (graduate students in psychology; psychology interns; practicum students; psychiatric residents)
- 3= parents
- 4= teacher
- 5= other non-professionals
- 6= experimenter
- 7= mixed
- 8= unknown

## V. Comparison Information

## \_\_\_\_\_ 21. Comparison number

## \_\_\_\_\_ 22. Type of comparison

- 1= treatment vs. control
- 2= behavioral vs. nonbehavioral
- 3= individual vs. group
- 4= combination

## \_\_\_\_\_ 23. Type of control group

- 1= none
- 2= no treatment
- 3= waiting list
- 4= attention-placebo
- 5= other
- 6= not available
- 7= mixed

## \_\_\_\_\_ 24. Sample size of treatment group (this comparison)

## \_\_\_\_\_ 25. Sample size of control group (this comparison)

## Appendix B (continued)

## VI. Treatment Characteristics

## \_\_\_\_\_ 26. Type of treatment

- 1= behavioral
- 2= nonbehavioral
- 3= mixed
- 4= unknown

## \_\_\_\_\_ 27. Method of delivery

- 1= individual
- 2= group
- 3= mixed
- 4= unknown

## \_\_\_\_\_ 28. Number of treatment sessions

## \_\_\_\_\_ 29. Average length of treatment sessions in minutes

## \_\_\_\_\_ 30. Treatment setting

- 1= school
- 2= home
- 3= mental health, community mental health, or psychology/psychiatry clinic
- 4= general hospital or dental clinic
- 5= residential treatment center (psychiatric or or special school)
- 6= camp
- 7= combination of at least two of the above
- 8= other
- 9= unknown

## VII. Characteristics of Outcome Measures

## \_\_\_\_\_ 31. Type of outcome measure

- 1= independent behavioral observation
- 2= nonindependent behavioral observation
- 3= peer sociometric
- 4= normed rating scale or behavioral checklist
- 5= non-normative/experimenter constructed instrument
- 6= achievement test or intellectual measure
- 7= other performance measure
- 8= school grades
- 9= objective performance measure (e.g. approaching

feared                                      object, dry nights)

## \_\_\_\_\_ 32. Impact of treatment measured

- 1= specific
- 2= generalized



## Appendix B (continued)

- \_\_\_\_\_ 33. Source of outcome measure
- 1= independent observers
  - 2= parents
  - 3= therapist
  - 4= teachers/school
  - 5= peers
  - 6= subject self-report
  - 7= subject performance measure
  - 8= other (e.g., expert judges, other than 1-7)
  - 9= mixed
  - 10= unknown
- \_\_\_\_\_ 34. Dimension of adjustment
- 1= fear/anxiety
  - 2= cognitive skills
  - 3= global adjustment
  - 4= social adjustment/social skills
  - 5= achievement
  - 6= personality
  - 7= self-esteem
  - 8= bed-wetting
  - 9= mixed
  - 10= unknown
  - 11= physiological
- \_\_\_\_\_ 35. Type of adjustment or change measured
- 1= behavioral
  - 2= personality
  - 3= academic performance
  - 4= sociometric
  - 5= cognitive tempo
  - 6= cognitive problem-solving skills
  - 7= physiological measure
  - 8= other
  - 9= mixed

## VIII. Effect Size Information

- \_\_\_\_\_ 36. Reliability of measure
- \_\_\_\_\_ 37. Effect size at posttreatment
- \_\_\_\_\_ 38. Length of follow-up in weeks
- \_\_\_\_\_ 39. Effect size at follow-up

## Appendix B (continued)

## \_\_\_\_\_ 40. How effect size was calculated

- 1= means and standard deviations
- 2= ANOVA summary table
- 3= t score
- 4= raw data
- 5= ANCOVA
- 6= probit or chi square/nonparametric
- 7= change scores
- 8= estimate from p
- 9= correlations
- 10= nonsignificant with no statistical information
- 11= Holmes method
- 12= posttest adjustment
- 13= mixed
- 14= separate methods for posttest and follow-up

## \_\_\_\_\_ 41. Source of data

- 1= standard information provided
- 2= data drawn from graphs
- 3= 2-week test-retest reliabilities used with change scores

## Appendix C

### Effect size calculations for special cases

#### Case I

For studies in which authors reported nonsignificant effects without presenting statistics, effect sizes were set at zero.

#### Case II

For studies in which means or standard deviations were not given, but an actual  $t$ -test was reported, the effect size was calculated as two times the value of  $t$  divided by the square root of the degrees of freedom.

#### Case III

For studies in which only an  $F$  value was given,  $t$  was calculated as the square root of " $F$ ." Then, the effect size was calculated from  $t$  as described above.

#### Case IV

Studies in which it was only reported that a  $t$ -test of  $F$ -test was significant at a specific level of  $p$  for a specific number of subjects, the corresponding  $t$  or  $F$  values were obtained from their respective tables. Effect sizes were then calculated from these values as described above.

#### Case V

For studies using a repeated measures design and analysis and a treatment versus control group design, the effect size was calculated from the value of interaction  $F$  as described above. If degrees of freedom were not reported, they were calculated as the number of subjects minus two. If the study failed to report the interaction  $F$  value, the  $F$  for groups was used.

#### Case VI

Studies comparing more than one treatment and/or control group and providing a post-hoc test of group differences required that these tests be converted into  $t$  values for effect size calculations. If the total number of subjects was given but the number per group was not, group sizes were assumed to be equal. If a  $p$  level was not reported, it was assumed to be 0.05. Using the number of subjects per group and the  $p$  level, a  $t$  value was determined by working from the  $t$  table. Then, the effect size was calculated as previously described.

#### Case VII

In cases where studies reported only frequencies or percentages, these statistics were transformed into  $z$ -scores by

## Appendix C (continued)

referring to the z-score table. The effect size was then calculated as the product of the experimental group's z-score minus the control group's z-score.

### Case VIII

For studies utilizing more than two groups wherein an  $F$  was computed, and the  $F$  summary table was provided along with the mean for each group, effect sizes were calculated as the mean of the treatment group minus the mean of the control group, divided by the square root of the mean square between groups divided by the  $F$  value for between groups.

### Case IX

Finally, for studies in which there were pre-treatment differences between groups, the effect size at PT was adjusted by subtracting the difference at pre-treatment.

## Appendix D

## Studies included in the meta-analysis

- Addesso, V.J., & Lipson, J.W. (1981). Group training of parents as therapists for their children. Behavior Therapy, 12, 625-633.
- Allen, R.P., Safer, D.J., Heaton, R., Ward, A., & Barrell, M. (1975). Behavior therapy for socially ineffective children. Journal of the American Academy of Child Psychiatry, 14(3), 500-510.
- Arnold, S.C., & Forehand, R. (1978). A comparison of cognitive training and response cost procedures in modifying cognitive styles of impulsive children. Cognitive Therapy and Research, 2(2), 183-187.
- Bandura, A., Grusec, J.E., & Menlove, F.L. (1967). Vicarious extinction of avoidance behavior. Journal of Personality and Social Psychology, 5, 16-23.
- Bandura, A., & Menlove, F.L. (1968). Factors determining vicarious extinction of avoidance behavior through symbolic modeling. Journal of Personality and Social Psychology, 8(2), 99-108.
- Christensen, A., Johnson, S.M., Phillips, S., & Glasgow, R.E. (1980). Cost effectiveness in behavioral family therapy. Behavior Therapy, 11, 208-226.
- Costantino, G., Malgady, R.G., & Rogler, L.H. (1986). Cuento Therapy: A culturally sensitive modality for Puerto Rican children. Journal of Consulting and

## Appendix D (continued)

Clinical Psychology, 54, 639-645.

Cullinan, D., Epstein, M.H., & Silver, L. (1977).

Modification of impulsive tempo in learning-disabled pupils. Journal of Abnormal Child Psychology, 5, 437-444.

Douglas, V.I., Parry, P., Marton, P. & Garson, C. (1976).

Assessment of a cognitive training program for hyperactive children. Journal of Abnormal Child Psychology, 4, 389-410.

Drummond, D.J. (1975). Self-instructional training: An approach to disruptive classroom behavior. Dissertation Abstracts International, 35(8-B), 4167-4168B.

Edleson, J.L., & Rose, S.D. (1981). Investigations into the efficacy of short-term group social skills training for socially isolated children. Child Behavior Therapy, 3, 1-16.

Egeland, B. (1974). Training impulsive children in the use of more efficient scanning techniques. Child Development, 45, 165-171.

Elias, M.J. (1983). Improving coping skills of emotionally disturbed boys through television-based social problem solving. American Journal of Orthopsychiatry, 53, 61-72.

Elliot, C. D. & Pumfrey, P. D. (1972). The effects of non-directive play therapy on some maladjusted boys.

## Appendix D (continued)

Education Research, 14, 157-161.

Flemming, C.C. (1982). Evaluation of an anger management program with aggressive children in residential treatment. (Doctoral Dissertation, Virginia Polytech Institute and State University, 1982). Dissertation Abstracts International, 4143b.

Garrison, S.R., & Stolberg, A.L. (1983). Modification of anger in children by affective imagery training. Journal of Abnormal Child Psychology, 11, 115-130.

Glenwick, D.S., & Barocas, R. (1979). Training impulsive children in verbal self-control by use of natural change agents. Journal of Special Education, 13, 387-398.

Goodman, G. (1972). Companionship therapy. San Francisco: Jossey-Bass.

Gottman, J. (1977). The effects of a modeling film on social isolation in preschool children: A methodological investigation. Journal of Abnormal Child Psychology, 5(1), 69-79.

Graves, T., Meyers, A.W., & Clark, L. (1988). An evaluation of parental problem-solving training in the behavioral treatment of childhood obesity. Journal of Consulting and Clinical Psychology, 56(2), 246-250.

Greenleaf, D. O. (1982). The use of structured learning therapy and transfer programming with disruptive adolescents in a school setting. Journal of School

## Appendix D (continued)

Psychology, 20, 122-130.

Gresham, F.M., & Nagle, R.J. (1980). Social skills training with children: Responsiveness to modeling and coaching as a function of peer orientation. Journal of Consulting and Clinical Psychology, 48, 718-729.

Halliwell, J.W., Musella, D.F., & Silvino, P.J. (1970). Effects of counseling on attitudes and grades with intermediate grade pupils designated as having poor attitudes. Elementary School Guidance and Counseling, 5, 113-123.

Hampe, E., Noble, H., Miller, L.C., & Barrett, C.L. (1973). Phobic children one and two years posttreatment. Journal of Abnormal Psychology, 82, 446-453.

Hansen, J.C., Niland, T.M., & Zani, L.P. (1969). Model reinforcement in group counseling with elementary school children. Personnel and Guidance Journal, 47, 741-744.

Heaton, R.C., Safer, D.J., Allen, R.P., Spinnato, Sr., N.C., & Prumo, F.M. (1976). A motivational environment for behaviorally deviant junior high school students. Journal of Abnormal Child Psychology, 4(3), 263-275.

Holmes, C.S. (1981). Reflective training and causal attributions in impulsive mildly retarded children. Journal of Clinical Child Psychology, 10, 194-199.

Jakibchuk, Z, & Smeriglio, V.L. (1976). The influence of symbolic modeling on the social behavior of preschool



## Appendix D (continued)

children with low levels of social responsiveness.

Child Development, 47, 838-841.

Kazdin, A.E., Esveltd-Dawson, K., French, N.H., & Unis, A.S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. Journal of Consulting and Clinical Psychology, 55, 76-85.

Kazdin, A.E., Esveltd-Dawson, K., French, N.H., & Unis, A.S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. Journal of Consulting and Clinical Psychology, 55, 76-85.

Keller, M.F., & Carlson, P.M. (1974). The use of symbolic modeling to promote social skills in preschool children with low levels of social responsiveness. Child Development, 45, 912-919.

Kendall, P.C., & Braswell, L. (1982). Cognitive-behavioral self-control therapy for children: A components analysis. Journal of Consulting and Clinical Psychology, 50, 672-689.

Kendall, P.C., & Finch, Jr., A.J. (1978). A cognitive-behavioral treatment for impulsivity: A group comparison study. Journal of Consulting and Clinical Psychology, 46, 110-118.

Kendall, P.C., & Wilcox, L.E. (1980). Cognitive-behavioral treatment for impulsivity: Concrete versus conceptual

## Appendix D (continued)

training in non-self-controlled problem children.

Journal of Consulting and Clinical Psychology, 48, 80-91.

Kendall, P.C., & Zupan, B.A. (1981). Individual versus group application of cognitive-behavioral self-control procedures with children. Behavior Therapy, 12, 344-359.

Kent, R.N., & O'Leary, K.D. (1976). A controlled evaluation of behavior modification with conduct problem children. Journal of Consulting and Clinical Psychology, 44, 586-596.

Kent, R.N., & O'Leary, K.D. (1977). Treatment of conduct problem children: BA and/or PhD therapists. Behavior Therapy, 8, 653-658.

Kirschenbaum, D.S., Harris, E.S., & Tomarken, A.J. (1984). Effects of parental involvement in behavioral weight loss therapy for preadolescents. Behavior Therapy, 15, 485-500.

Kolvin, I., Garside, R.F., Nicol, A.R., MacMillan, A., Wolstenholme, F., & Leitch, I.M. (1981). Help starts here: The maladjusted child in the ordinary school. New York: Tavistock.

Kranzler, G.D., Mayer, G.R., Dyer, C.O., & Munger, P.F. (1966). Counseling with elementary school children: An experimental study. Personnel and Guidance Journal, 44, 944-949.

Krop, H., Calhoun, B., & Verrier, R. (1971). Modification

## Appendix D (continued)

- of the "self-concept" of emotionally disturbed children by covert reinforcement. Behavior Therapy, 2, 201-204.
- Labbe, E., & Williamson, D. (1984). Treatment of childhood migraine using autogenic feedback training. Journal of Consulting and Clinical Psychology, 52, 968-976.
- Ladd, G.W. (1981). Effectiveness of a social learning method for enhancing children's social interaction and peer acceptance. Child Development, 52, 171-178.
- Lebovitz, C.K. (1983). Filial therapy: Outcome and process. (Doctoral dissertation, Texas Tech University, 1982). Dissertation Abstracts International, 43(12-B), 4152.
- Lilly, M.S. (1971). Improving social acceptance of low sociometric status, low achieving students. Exceptional Children, 37, 341-347.
- Litrownik, A.J., Freitas, J.L., & Franzini, L.R. (1978). Self-regulation in mentally retarded children: Assessment and training of self-monitoring skills. American Journal of Mental Deficiency, 82(5), 499-506.
- Lott, S.J.B. (1975). A study of covert reinforcement as a treatment of test anxiety in male and female fifth graders. Dissertation Abstracts International, 36(4-A), 2028A.
- McGillivray, J.A., Cummins, R.A., & Prior, M.R. (1988). Cognitive behaviour modification of impulsive responding

## Appendix D (continued)

by hyperaggressive children in interpersonal problem situations. Australia and New Zealand Journal of Development Disabilities, 14(1), 55-70.

Michenbaum, D.H., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. Journal of Abnormal Psychology, 77, 115-126.

Michelson, L., Mannarino, A.P., Marchione, K.E., Stern, M., Figueroa, J., & Beck, S. (1983). A comparative outcome study of behavioral social-skills training, interpersonal problem-solving and non-directive control treatments with child psychiatric outpatients. Behavior Research and Therapy, 21(5), 545-556.

Miller, L. C., Barrett, C.L., Hampe, E., & Noble, H. (1972). Comparison of reciprocal inhibition, psychotherapy, and waiting list control for phobic children. Journal of Abnormal Psychology, 79(3), 269-279.

Nagle, R.J., & Thwaite, B.C. (1979). Modeling effects on impulsivity with learning disabled children. Journal of Learning Disabilities, 12, 51-56.

Oden, S., & Asher, S.R. (1977). Coaching children in social skills for friendship making. Child Development, 48, 495-506.

Palkes, H., Stewart, M., & Freedman, J. (1972). Improvement in maze performance of hyperactive boys as a function

## Appendix D (continued)

- of verbal-training procedures. The Journal of Special Education, 5, 337-342.
- Parrish, J.M., & Erickson, M.T. (1981). A comparison of cognitive strategies in modifying the cognitive style of impulsive third-grade children. Cognitive Therapy and Research, 5, 71-84.
- Pihl, R.O., Parkes, M., Drake, H., & Vrana, F. (1980). The intervention of a modulator with learning disabled children. Journal of Clinical Psychology, 36(4), 972.
- Pisterman, S., McGrath, P., Firestone, P., Goodman, J.T., Webster, I., and Mallory, R. (1989). Outcome of parent-mediated treatment of preschoolers with attention deficit disorder with hyperactivity. Journal of Consulting and Clinical Psychology, 57, 628-635.
- Reid, M.K., & Borkowski, J.G. (1987). Causal attributions in hyperactive children: Implications for teaching strategies and self-control. Journal of Educational Psychology, 79, 296-307.
- Richter, I.L., McGrath, P.J., Humphreys, P.J., Goodman, J.T., Firestone, P., & Keene, D. (1986). Cognitive and relaxation treatment of paediatric migraine. Pain, 25, 195-203.
- Rickel, A.U., Eshelman, A.K., & Loigman, G.A. (1983). Social problem solving training: A follow-up study of cognitive and behavioral effects. Journal of Abnormal

## Appendix D (continued)

Child Psychology, 11, 15-28.

Rickel, A.U., Smith, R.L., & Sharp, K.C. (1979).

Description and evaluation of a preventive mental health program for preschoolers. Journal of Abnormal Child Psychology, 7, 101-112.

Ridberg, E.H., Parke, R.D., & Hetherington, E.M. (1971).

Modification of impulsive and reflective cognitive styles through observation of film-mediated models.

Developmental Psychology, 5, 369-371.

Rotheram, M.J. (1982). Social skills training with under achievers, disruptive, and exceptional children.

Psychology in the Schools, 19, 532-539.

Rucker, C.N., & Vincenzo, F.M. (1970). Maintaining social acceptance gains made by mentally retarded children.

Exceptional Children, 36, 679-680.

Safer, D.J., Heaton, R.C., & Parker, F.C. (1981). A

behavioral program for disruptive junior high school students: Results and follow-up. Journal of Abnormal Child Psychology, 9, 483-494.

Sage, C.K. (1980). Using self-instructional techniques to teach children positive assertion. Dissertation

Abstracts International, 41(1-B), 367.

Saylor, C.F., Benson, B., & Einhaus, L. (1985). Evaluation of an anger management program for aggressive boys in inpatient treatment. Journal of Child and Adolescent

## Appendix D (continued)

Psychotherapy, 2, 5-16.

Scovern, A.W., Bukstel, L.H., Kilmann, P.R., Laval, R.A., Busemeyer, J., & Smith, V. (1980). Effects of parent counseling on the family system. Journal of Counseling Psychology, 27, 268-275.

Schumacher, R.R. (1974). The influence of differential group composition on the effectiveness of group counseling with second, third, fourth, and fifth grade male children. (Doctoral dissertation, University of Virginia, 1974). Dissertation Abstracts International, 35(4-A), 1992-1993.

Seeman, J., Barry, E., & Ellingwood, C. (1964).

Interpersonal assessment of play therapy outcomes.

Psychotherapy: Theory, Research, and Practice, 1, 64-66.

Senediak, C., & Spence, S.H. (1985). Rapid versus gradual scheduling of therapeutic contact in a family based behavioural weight control programme for children. Behavioural Psychotherapy, 13, 265-287.

Siegert, F.E., & Yates, B.T. (1980). Behavioral child-management cost-effectiveness: A comparisonn of individual in- office, individual in-home, and group delivery systems. Evaluation and the Health Professions, 3, 123-152.

Soper, E.T. (1978). The effects of a "think aloud" program on locus of control in young boys. (Doctoral

## Appendix D (continued)

- dissertation, University of Utah, 1977). Dissertation Abstracts International, 38(7-A), 3966.
- Tefft, B.M., & Kloba, J.A. (1981). Underachieving high school students as mental health aides with maladapting primary-grade children. American Journal of Community Psychology, 9, 303-319.
- Temerlin, M.K., Trousdale, W.W., LaCrone, H.H., Harrison, C.H., Rundell, O.H. (1967). Effects of increased mothering and skin contact with retarded boys. American Journal of Mental Deficiency, 71, 890-893.
- Tosi, D.J., Swanson, C., & McLean, P. (1969). Group counseling with nonverbalizing elementary school children. Elementary School Guidance and Counseling, 4, 260-266.
- Weinrott, M.R., Corson, J.A., & Wilchesky, M. (1979). Teacher-mediated treatment of social withdrawal. Behavior Therapy, 10, 281-294.
- Weissbrod, C.S., & Bryan, J.H. (1973). Filmed treatment as an effective fear-reduction technique. Journal of Abnormal Child Psychology, 1, 196-201.
- White, W.C., & Davis, M.T. (1974). Vicarious extinction of phobic behavior in early childhood. Journal of Abnormal Child Psychology, 2, 25-32.
- Wilson, N.H., & Rotter, J.C. (1986). Anxiety management training and study skills counseling for students on



Appendix D (continued)

self-esteem and test anxiety performance. The School Counselor, 34, 18.

Wolfson, J. (1982). Self-talk and exposure in the reduction of avoidance in aquaphobic children. Dissertation Abstracts International, 43, 2721B.

Woollacott, S., Graham, P., & Stevenson, J. (1978). A controlled evaluation of the therapeutic effectiveness of a psychiatric day centre for pre-school children. British Journal of Psychiatry, 132, 349-355.

## References

- Achenbach, T.M. (1978). Psychopathology of childhood: Research problems and issues. Journal of Consulting and Clinical Psychology, 46, 759-776.
- Achenbach, T.M., & Edelbrock, C.S. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont Press.
- Barnett, R.J., Docherty, J.P., & Frommelt, G.M. (1991). A review of child psychotherapy research since 1963. Journal of the American Academy of Child and Adolescent Psychiatry, 30, 1-14.
- Bryant, F.B. (1986). Improving the quality of research synthesis in program evaluation. Policy Studies Review, 5, 708-721.
- Casey, R.J., & Berman, J.S. (1985). The outcome of psychotherapy with children. Psychological Bulletin, 98, 388-400.
- Dorfman, E. (1958). Personality outcome of client-centered child therapy. Psychological Monographs, 72, 456.
- Durlak, J.A. (1979). The comparative effectiveness of paraprofessional and professional helpers. Psychological Bulletin, 86, 80-92.
- Durlak, J.A., Fuhrman, T., & Lampman, C. (1991). Effectiveness of cognitive-behavior therapy for

maladapting children: A meta-analysis. Psychological Bulletin, 110, 204-214.

Durlak, J.A., Lampman, C., and Wells, A. (1990). The effectiveness of different forms of child psychotherapy. A paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 3, 1990.

Gottman, J. (1977). The effects of a modeling film on social isolation in preschool children: A methodological investigation. Journal of Abnormal Child Psychology, 5, 69-79.

Hampe, E., Noble, H., Miller, L.C., & Barrett, C.L. (1973). Phobic children one and two years posttreatment. Journal of Abnormal Psychology, 82, 446-453.

Hedges, L.V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, Fl.: Academic Press.

Heinicke, C.M. (1969). Frequency of psychotherapeutic session as a factor affecting outcome: Analysis of clinical ratings and test results. Journal of Abnormal Psychology, 74, 553-560.

Jakibchuk, Z, & Smeriglio, V.L. (1976). The influence of symbolic modeling on the social behavior of preschool children with low levels of social responsiveness. Child Development, 47, 838-841.

Kazdin, A.E. (1993). Psychotherapy for children and adolescents: Current progress and future directions.

American Psychologist, 48, 644-687.

Kazdin, A.E., Bass, D., Ayers, W.A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. Journal of Consulting and Clinical Psychology, 58, 729-740.

Kazdin, A.E., Siegel, T.C., & Bass, D. (1990). Drawing on clinical practice to inform research on child and adolescent psychotherapy: survey of practitioners. Professional Psychology: Research and Practice, 3, 189-198.

Keller, M.F., & Carlson, P.M. (1974). The use of symbolic modeling to promote social skills in preschool children with low levels of social responsiveness. Child Development, 45, 912-919.

Kendall, P.C., & Norton-Ford, J.D. (1982). Therapy outcome research methods. In P. C. Kendall & J. N. Butcher (Eds.), Handbook of Research Methods in Clinical Psychology (pp. 429-460). New York: Wiley.

Lehrman, L.J., Sirluck, H., Black, B.J., & Glick, S.J. (1949). Success and failure of treatment of children in the child guidance clinics of the Jewish Board of Guardians. Jewish Board of Guardians Research Monographs, 1, 1-87.

Love, L.R., Kaswan, J., & Bugental, D.E. (1972). Differential effectiveness of three clinical interventions for different socioeconomic groupings.

Journal of Consulting and Clinical Psychology, 39, 347-360.

McGillivray, J.A., Cummins, R.A., & Prior, M.R. (1988).

Cognitive behaviour modification of impulsive responding by hyperaggressive children in interpersonal problem situations. Australia and New Zealand Journal of Development Disabilities, 14(1), 55-70.

Miller, L.C., Barrett, C.L., Hampe, E., & Noble, H. (1972).

Comparison of reciprocal inhibition, psychotherapy, and waiting list control for phobic children. Journal of Abnormal Psychology, 79, 269-279.

Nicholson, R.A., & Berman, J.S. (1983). Is follow-up

necessary in evaluating psychotherapy? Psychological Bulletin, 93, 261-278.

Nietzel, M.T., & Fisher, S.G. (1981). Effectiveness of

professional and paraprofessional helpers: A comment on Durlak. Psychological Bulletin, 89, 555-565.

Orlinsky, D.E., & Howard, K.I. (1986). Process and outcome

in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), Handbook of Psychotherapy and Behavior Change (3rd ed., pp. 311-381). New York: Wiley.

Raimy, V. (1952). Clinical methods: Psychotherapy. Annual

Review of Psychology, 3, 321-350.

Robins, L., & Rutter, M. (Ed.). (1990). Straight and devious

pathways from childhood to adulthood. Cambridge, England: Cambridge University Press.

- Rutter, M. (1984). Psychopathology and development: Childhood antecedents of adult psychiatric disorder. Australian and New Zealand Journal of Psychiatry, 18, 225-234.
- Rutter, M. (1985). Family and school influences on behavioral development. Journal of Child Psychology and Psychiatry, 26, 349-368.
- Seeman, J., Barry, E., & Ellinwood, C. (1964). Interpersonal assessment of play therapy outcome. Psychotherapy: Theory, Research, and Practice, 1, 64-66.
- Shapiro, D.A., Harper, H., Startup, M., Reynolds, S., Bird, D., & Soukas, A. (1994). The high-water mark of the drug metaphor: A meta-analytic critique of process-outcome research. In R. L. Russell (Ed.), Reassessing Psychotherapy Research, (pp. 1-36). New York: Guilford Press.
- Smith, M.L. (1980). Publication bias in meta-analysis. Evaluation in Education, 4, 22-24.
- Smith, M.L., Glass, G.V., & Miller, T.I. (1980). The benefits of psychotherapy. Baltimore, Md.: Johns Hopkins University Press.
- Stiles, W.B., Shapiro, D.A., and Harper, H. (1994). Finding the way from process to outcome: Blind alleys and unmarked trails. In R. L. Russell (Ed.), Reassessing Psychotherapy Research (pp. 36-64). New York: Guilford Press.

- Weinrott, M.R., Corson, J.A., & Wilchesky, M. (1979).  
Teacher-mediated treatment of social withdrawal.  
Behavior Therapy, 10, 281-294.
- Weissbrod, C.S., & Bryan, J.H. (1973). Filmed treatment as  
an effective fear-reduction technique. Journal of  
Abnormal Child Psychology, 1, 196-201.
- Weisz, J.R., Weiss, B., Alicke, M.D., & Klotz, M.L. (1987).  
Effectiveness of psychotherapy with children and  
adolescents: Meta-analytic findings for clinicians.  
Journal of Consulting and Clinical Psychology, 55, 542-  
549.
- Weisz, J.R., Weiss, B., & Donenberg, G.R. (1992). The lab  
versus the clinic: Effects of child and adolescent  
psychotherapy. American Psychologist, 47, 1578-1585.
- Weiss, G., & Hechtman, L.T. (1986). Hyperactive children  
grown up. New York: Guilford Press.
- Wolff (1986).
- Wright, D.M., Moelis, I., & Pollack, L.J. (1976). The  
outcome of individual child psychotherapy: Increments at  
follow-up. Journal of Child Psychology and Psychiatry,  
17, 275-285.

## VITA

James Kelly Cotten, son of Lutie Cotten Darnell of Germantown, Tennessee, was born in Lubbock, Texas on November 10, 1967.

Mr. Cotten began attending Memphis State University in Memphis, Tennessee in August, 1986. He graduated cum laude with a Bachelor of Arts degree from Memphis State University in May, 1992, double majoring in Psychology and Philosophy.

He began graduate school at Loyola University of Chicago in August, 1992 in the Clinical Division of the Department of Psychology. At present, Mr. Cotten is pursuing the Doctor of Philosophy degree at Loyola. In addition to his studies, Mr. Cotten is actively engaged in scholarly research for presentation and publication.



## THESIS APPROVAL SHEET

The thesis, "Do the Effects of Child Psychotherapy Vary Over Time?" submitted by James Kelly Cotten has been read and approved by the following committee:

Dr. Joseph Durlak, Associate Professor  
Department of Psychology, Loyola University of Chicago

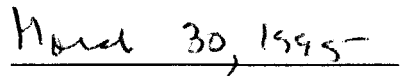
Dr. Fred Bryant, Associate Professor  
Department of Psychology, Loyola University of Chicago

The final copies have been examined by the director of the thesis committee and the signature which appears below verifies the fact that any necessary changes have been incorporated and the the thesis is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of master of arts.



Director's Signature



Date